Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 11, No. 3, 2025

Prediction of the Remaining Life of Rolling Bearings based on the Classification of Degradation States

Bingxu Zhou *, Yang Yu

School of Information Science and Engineering, Shenyang University of Technology, Shenyang Liaoning, 110870, China
* **Corresponding author**: Bingxu Zhou (Email: 1584723065@qq.com)

Abstract: view of the fact that the healthy running time of rolling bearings in actual working conditions lasts for a long time, and the vibration signal fluctuation in the healthy operation stage is relatively stable, and it is difficult to extract useful degradation information from them, a remaining life prediction model based on the classification of degradation states based on the hidden Markov model is proposed. Firstly, the characteristics of the bearing vibration signal are extracted and the dimension reduction is carried out, and then the RMS is used as the observation sequence to divide the degradation state. Finally, the LightGBM optimized by the whale algorithm is used to predict the remaining life of the degraded stage after division. The experimental data set of XJTU-XY bearing is used for verification, and the results show that the prediction performance of LightGBM after whale algorithm optimization is significantly improved.

Keywords: Degenerate State Classification; Hidden Markov Model; Whale Optimization Algorithm; LightGBM.

1. Introduction

The entire cycle of the vibration signal of a rolling bearing consists of both a healthy and a degraded phase [1]. In the healthy operation phase, the characteristics of the bearing change slowly, but in the degradation stage, the characteristics change rapidly and nonlinearly. The state segmentation helps the model focus on the key degradation stages and optimize the prediction effect. It not only improves the accuracy and robustness of the model, but also provides early warning capabilities that can be better adapted to real-world industrial application scenarios [2].

There are generally two types of health status classification methods: quantitative and qualitative [3]. In Ref. [4], an adaptive state division method is adopted, the sliding window is used to continuously update the early warning range of the 3σ criterion, and the continuous trigger mechanism is combined to adaptively determine the degradation start time for state division. In Ref. [5], a state threshold is set based on Chebyshev's inequality, which is used to determine when a device enters a degraded state. The above quantitative classification methods are often too simplistic and not suitable for more complex systems. Ref. [6] uses the grey clustering method to divide the health status of ships. This qualitative method of clustering-based state segmentation can often only start from data points, and the result is usually only the health state corresponding to each data point, without considering the impact on subsequent moments.

In view of the shortcomings and deficiencies of the above state division methods, this paper proposes a state division method for rolling bearings based on the hidden Markov model, and combines the remaining life prediction with LightGBM optimized by whale algorithm. Because the hidden Markov model can effectively train and decode the time series data, obtain the hidden state sequence corresponding to each state to realize the state division, which provides a good basis for the prediction of the subsequent remaining life.

2. Rationale

2.1. Hidden Markov Model

The hidden Markov model is a statistical model widely used for time series analysis, which is mainly composed of observation sequences, initial state distributions, state transition probabilities and observed probabilities, which are used to describe the sequences generated by the implicit Markov process.

(1) Observation sequence

Contains all possible observations with the expression

$$O = \{o_1, o_2, \dots, o_T\} \tag{1}$$

(2) Initial state probability distribution π_i

Indicates the probability that the system belongs to a state s_i at the initial moment:

$$\pi_i = P(S_1 = s_i), i = 1,2, \dots, N \tag{2}$$

(3) State transition probability matrix A

Represents the probability that a system will move from one state to another:

$$\mathrm{A} = \{a_{ij}\}, a_{ij} = P\left(S_{t+1} = s_j | S_t = s_i\right) \tag{3}$$

Satisfy:

$$\sum_{j=1}^{N} a_{ij} = 1, \forall i \tag{4}$$

(4) Observed probability distribution B

Define the probability of generating an observation in a hidden state:

$$\mathrm{B} = \left\{b_j(k)\right\}, b_j(k) = P\left(O_t = o_k \middle| S_t = s_j\right) \tag{5}$$

satisfy

$$\sum_{k=1}^{M} b_j(k) = 1, \forall j \tag{6}$$

2.2. Whale Optimization Algorithms

In 2016, Seyedali Mirjalili first proposed a whale optimization algorithm inspired by mimicking the hunting behavior of humpback whales. There are two main methods used in the humpback whale hunt: encirclement and bubble net. During the hunt, the humpback whale's choice of these

two methods is random. In addition, each individual whale is faced with a choice: whether to swim to the individual whale closest to them or to the individual whale in the optimal position. The whale algorithm is mainly divided into the following steps:

First, the position of the whale in the WOA search space is randomly generated by some search particles, when $|A| < 1$, WOA enters the local search, and when$|A| > 1$, WOA enters the global search.

$$a(t) = 2 - \frac{2t}{T} \quad (7)$$

$$A(t) = 2a(t)r - a(t) \quad (8)$$

where t is the number of times it has been completed, T is the upper line of the number of iterations, and $r \in [0,1], A \in [-a, a]$.

(1) Local search stage

There are two main ways to search in the local search phase: shrink wrap and spiral update.

a) Shrinkage enveloping phase

At the beginning of the contraction envelope, A and a are positively correlated and gradually decrease. At the same time, the search particles are constantly moving to random locations in the global range, and in this way the search particles are constantly updated. Its updated position is shown in equation (9).

$$\vec{X}(t+1) = \overrightarrow{X^*}(t) - A(t) \cdot \vec{D}(t) \quad (9)$$

$$\vec{D}(t) = \left|C(t) \cdot \overrightarrow{X^*}(t) - \vec{X}(t)\right| \quad (10)$$

The gap between the search particle and the target can be calculated by equation (10) and further pursuit.

b) Spiral update phase

After the above update is completed, the corresponding new positions need to be given by the helical equation, as shown in equations (11) and (12).

$$\vec{X}(t+1) = \overrightarrow{D'}(t) \cdot e^{bl} \cdot \cos(2\pi l) + \overrightarrow{X^*}(t) \quad (11)$$

$$\overrightarrow{D'}(t) = \left|\overrightarrow{X*}(t) - \vec{X}(t)\right| \quad (12)$$

where $\overrightarrow{D'}(t)l$ is the difference between the search particle and the optimal solution, b is the constant coefficient, and l is the random number in the range [-1,1].

The random predation mode of whales can be expressed by $p \in [0,1]$, and the relationship between the position update mode of whales and the size of p-value is shown in equation (13), where when $p < 0.5$, the whale's position is updated by contraction and encirclement; When $p \geq 0.5$, the whale's position is updated in a spiral update.

$$\vec{X}(t+1) = \begin{cases} \overrightarrow{X^*}(t) - A(t) \cdot \vec{D}(t) & ,p < 0.5 \\ \overrightarrow{D'}(t) \cdot e^{bl} \cdot \cos(2\pi l) + \overrightarrow{X^*}(t) & ,p \geq 0.5 \end{cases} \quad (13)$$

(2) Global search stage

In the global search phase, the whale will use a random prey to update its location, while other whales will conduct a global search at a certain distance from the whale to achieve a comprehensive capture of the prey. The formula is shown in (14) and (15).

$$\vec{X}(t+1) = \overrightarrow{X_{rand}} - A(t) \cdot \vec{D}(t) \quad (14)$$

$$\vec{D}(t) = \left|C(t) \cdot \overrightarrow{X_{rand}} - \vec{X}(t)\right| \quad (15)$$

where $\overrightarrow{X_{rand}}$ is a randomly selected search particle.

X^* is updated by comparing it to the optimal particle, and the optimal value is continuously approximated until the optimal value is reached.

3. Experimental Datasets

3.1. Experimental Datasets

In this study, the life-cycle vibration signal data of rolling bearings (XJTU-SY) collected by Xi'an Jiaotong University was used for experimental verification. The bearing accelerated life test bench and its structural composition of the collected data are shown in Figure 2, and the test collected the data set of five bearings under three working conditions, with a sampling frequency of 25.6kHz and a sampling interval of 1min. Two bearing datasets for two of these working conditions were used for experiments, and their dataset information is shown in Table 1.

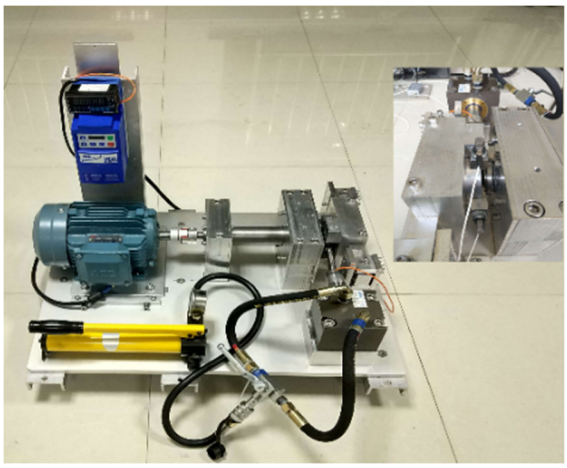


Fig 1. Bearing testbed

Table 1. Information about some XJTU-SY datasets

Operating conditions	Data set	Number of samples	Actual lifespan
1	Bearing1_1	123	2h3min
2	Bearing2_1	491	8h11min

3.2. Feature extraction and selection

In order to more comprehensively characterize the degradation state information of rolling bearings, and the vibration signals will reflect different degradation characteristics in different domains, the original vibration signal data of bearing 1_1 is divided into 3479 samples for feature extraction in the time domain, frequency domain and time frequency domain.

Table 2. Time-domain characteristics

Number	Characteristic	Number	Characteristic
1	peak	8	Skewness
2	average value	9	Kurtosis
3	Peak-to-peak	10	Waveform fact or
4	Absolute mean	11	Crest factor
5	standard deviation	12	Pulse factor
6	variance	13	Margin factor
7	root mean square		

3.2.1. Time-domain Feature Extraction

The time-domain characteristics of the vibration signal are

often a direct reflection of the operating state of the equipment. Time-domain analysis can help to determine whether the equipment is running smoothly or has potential failures. Table 2 shows the time domain features extracted in this document.

Table 3. Frequency domain characteristics

Number	Characteristic	Number	Characteristic
1	mean	7	Maximum amplitude
2	Frequency of the center of gravity	8	Root variance amplitude
3	Average frequency	9	Average phase angle
4	Root mean square frequency	10	Maximum phase angle
5	Root variance frequency	11	energy
6	Power spectral density	12	Extreme phase angles

3.2.2. Frequency Domain Feature Extraction

Frequency-domain feature extraction is also a key step in vibration signal analysis, which mainly reveals the frequency component and energy distribution in the signal, which can help to identify and distinguish the degree of bearing failure damage. Table 3 shows the frequency domain characteristics extracted by the fast Fourier transform

3.2.3. Time-frequency Domain Feature Extraction

Wavelet packet decomposition divides the signal into multiple frequency bands and analyzes the energy characteristics of each frequency band separately, and extracts the time-frequency domain characteristics of the corresponding frequency band in this way. In this paper, the three-layer wavelet decomposition of the bearing vibration signal is carried out by db3 wavelet, and the sub-signals of eight frequency bands are obtained by decomposition. Finally, the wavelet packet coefficients are calculated, and the energy characteristics of each frequency band are the eight time-frequency domain features extracted.

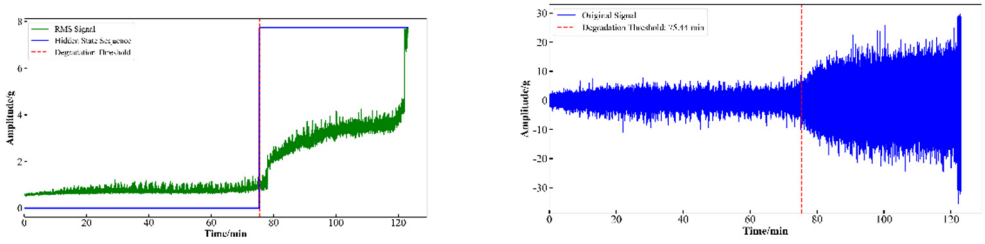
3.2.4. Feature Selection

In this paper, principal component analysis (PCA) is us ed to reduce the dimensionality of the extracted high-dim ensional feature set, and the cumulative variance contribut ion rate of 95% is selected as the dimensionality reductio n criterion, and the features of 33 dimensions are reduced to 15 dimensions.

3.3. State Division

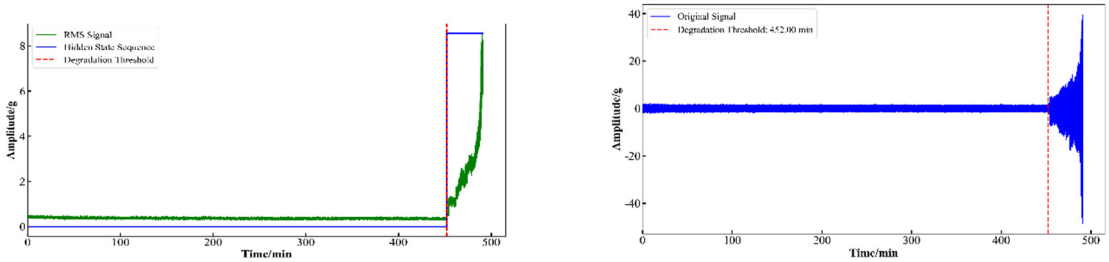
In order to verify the effectiveness of the partition effect of the hidden Markov model, the bearings of the test bearing 1_1 and the bearing 2_1 were divided, and the HMM was trained and decoded with their RMS as the observation sequence, because only two stages need to be divided: the healthy stage and the degenerative stage, so the number of hidden states was set to 2 to divide.

Fig. 3 and Fig. 4 are the results of the state division of bearing 1_1 and bearing 2_1 respectively, and their figure a is the hidden state sequence generated according to the observation sequence RMS, and the RMS value of the state division point changes significantly, indicating that the bearing begins to enter the degradation stage.



a RMS and hidden state sequences b Raw signals and their divisions

Fig 2. Bearing 1_1 division result



a RMS and hidden state sequences b Raw signals and their divisions

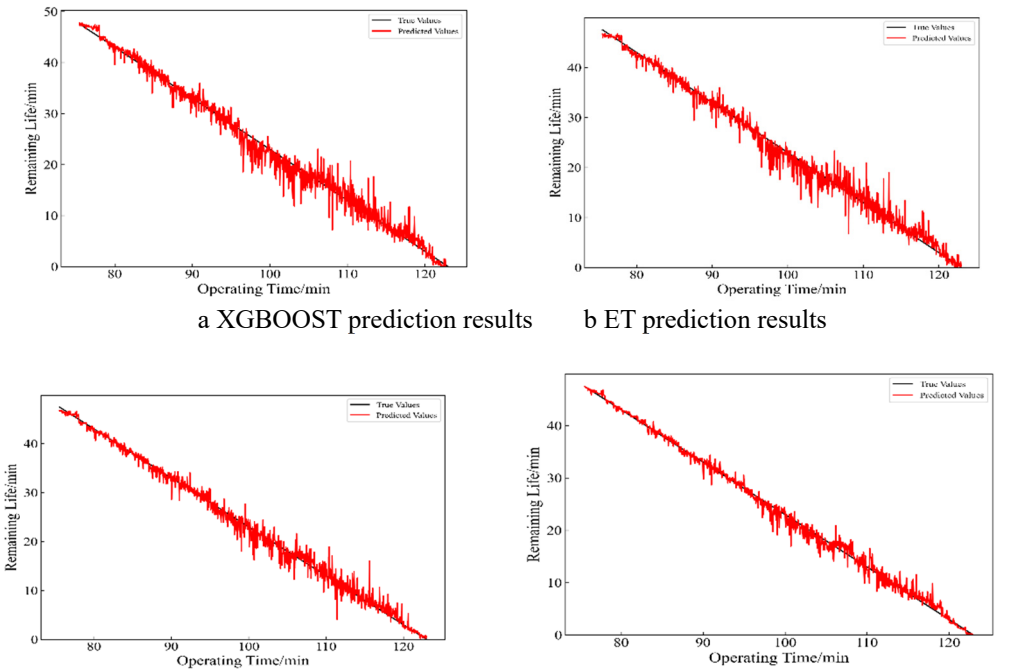
Fig 3. Bearing 2_1 division result

Fig. B is the result of dividing the original signal according to the degradation threshold point of Figure A, the original signal change before the degradation threshold point of the two bearings is relatively stable, and the signal fluctuates significantly from the degradation threshold point. According to the results of the RMS value of the two bearings, the generated hidden state sequence and the division of the original signal by the degradation threshold point, the hidden Markov model has a good effect on the division of the healthy phase and the degradation stage of the bearing operation, and is suitable for bearings under different working conditions.

3.4. Remaining Life Prediction

The principal component characteristics of the degradation

stage after bearing 1_1 division are used for life prediction, and the training set and the test set are divided into seven to three. The prediction performance of LightGBM optimized by whale algorithm is compared with LightGBM, Extreme Random Tree (ET) and XGBOOST without optimization, the results are shown in Figure 5, and finally the prediction effect is evaluated by R2 and RMSE, and the comparison results are shown in Table 4



a XGBOOST prediction results b ET prediction results

c LightGBM prediction results d WOA-LightGBM prediction results

Fig 4. Comparison chart of remaining life predictions

Table 4. Evaluation index scores

Predictive models	R2	RMSE
XGBOOST	0.9850	1.3451
ET	0.9865	1.2894
LightGBM	0.9912	0.8792
WOA-LightGBM	0.9948	0.5749

From the comparison results, it can be seen that the R2 and RMSE of the LightGBM model are lower than those of ET and XGBOOST, while the whale algorithm is optimized. The forecast R2 of LightGBM will increase by 0.0036, RMSE The decrease is 0.3043, which indicates that the prediction model has the best prediction performance and reflects the effectiveness of the model.

4. Conclusion

(1) The hidden Markov model can effectively process the time series data, and realize the accurate state division of the bearings under the two working conditions by combining the Baum-Welch algorithm and the Viterbi algorithm, and can effectively divide the whole life cycle of the bearing into the health stage and the degradation stage.

(2) Through the verification of the XJTU-SY dataset, the predictability of LightGBM is better than that of ET and XGBOOST, and the prediction performance of LightGBM after whale algorithm optimization has been significantly improved, which verifies the effectiveness of the model

References

[1] Han C, Xianguang K, Qibin W, et al.The two-stage RUL prediction across operation conditions using deep transfer learning and insufficient degradation data[J].Reliability Engineering and System Safety,2022,225.

[2] SHAO Haidong,XIAO Yiming,YAN Shen. Simulation data-driven improved unsupervised domain-adaptive bearing fault diagnosis[J]. Journal of Mechanical Engineering, 2023,59 (03): 76-85.

[3] Qiu LJ,Wu MH.A review of PHM technology framework and its key technologies[J]. Foreign Electronic Measurement Technology,2018,37(02):10-15.

[4] C.H. Hu,X.G. Wang,C.C. Wang,C.C. Wang,et al. Remaining service life prediction model based on state partitioning and integrated learning[J/OL]. Electromechanical Engineering,1-9[2025-03-05].

[5] Shakya P, Kulkarni S M, Darpe K A.A novel methodology for online detection of bearing health status for naturally progressing defect[J].Journal of Sound and Vibration,2014, 333 (21):5614-5629.

[6] Lv Jianwei, Yu Peng, Wei Jun, et al. Health state assessment method for naval equipment[J]. Journal of Naval Engineering University,2011,23(03):72-76.